IJMRSET

MULTIDISCIPLINARY · SCIENCE · ENGINEERING · TECHNOLOGY

e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

ISSN
INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.521

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707156 |

Detection of Phishing Websites Using Machine Learning

Karthik B S*[1], Praveen K S*[2]

Student, Master of Computer Applications, East West Institute of Technology, Bengaluru, Karnataka, India*[1]

Associate Professor, Master of Computer Applications, East West Institute of Technology, Bengaluru, Karnataka, India*[2]

ABSTRACT: Criminals who want to steal sensitive information make fake websites and email accounts. The email will include real company slogans and logos. The hacker grow admission to all of the user's confidential in sequence, counting depository description in rank, individual login passwords, and images, when the user clicks on a association provide by them. The accuracy of current systems' use of RF and DT algorithms must be improved. The latency of the current sculpts is low. There is no particular user interface for the current systems. Different algorithms are not compared in the current system. When the links or emails are opened, customers are taken to a fake website that looks like the real business. base on URL consequence skin texture, the reproduction are used to classify malware website and to select and implement the best device wisdom sculpt

I. INTRODUCTION

Phishing is a common method used to mislead unsuspecting people into providing personal information by utilising phoney websites. Phishing website URLs are intended to steal personal information such as user names, passwords, and online financial transactions. Phishers use websites that are visually and linguistically similar to legitimate websites. To prevent the rapid evolution of phishing techniques as a result of growing technology, anti-phishing approaches must be used to spot phishing. Machine learning is an effective method for preventing phishing attacks. Hackers typically use phishing because it is easier to trick a victim into opening a malicious link that appears to be legitimate than it is to try to circumvent a computer's security mechanisms. The malicious links inside the message body are designed to seem to lead to the faked firm by using its logos and other authentic information. Machine learning is applied in the method provided to build a breakthrough way for detecting phishing websites.

Key words: Phishing Detection, Machine Learning Models, Feature Extraction, Data Mining Techniques, Classification Algorithms, Behavioral Analysis, Feature Selection, Website Authentication, Supervised Learning, Unsupervised Learning

II. LITERATURE SURVEY

H. Huang et al., (2009) proposed the frameworks that distinguish the malware utilizing page section similitude that breaks down universal resource locator tokens to create forecast preciseness malware pages normally keep its CSS vogue like their objective pages.

S. Marchal et al., (2017) proposed this technique to differentiate Malware website depends on the examination of authentic site server log knowledge. An application Off-the- Hook application or identification of malware website. Free, displays a couple of outstanding properties together with high preciseness, whole autonomy, and nice language-freedom, speed of selection, flexibility to dynamic phish and flexibility to advancement in malware ways.

We have residential ours mission via a website as a platform for all the users. This is an interactive and responsive website that will be worn to detect whether a website is legitimate or malware. This website is made using different web designing languages which include HTML, CSS, Javascript and Flask framework in Python. The basic structure of the website is made with the help of HTML. CSS is used to add effects to the website and make it more attractive and user-friendly. It must be famous that the Since the website is intended for all users, it must be simple to use and should not present any difficulties for any user.

The proposed system is trained with the dataset consists of different features and note that the dataset don’t contain any website URL. The dataset consists of different features that are to be taken hooked on deliberation whilst influential a website URL as legitimate or malware.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707156 |

III. SYSTEM DESIGN

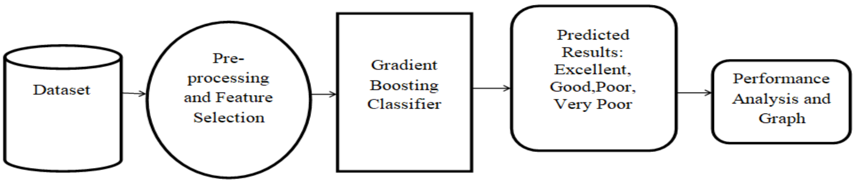

Dataset
Pre-processing and Feature Selection
Gradient Boosting Classifier
Predicted Results: Excellent, Good,Poor, Very Poor
Performance Analysis and Graph

Figure 1: System Architecture

IV. RESULTS AND OUTCOMES

Dataset Collection and Preparation:

- Gathering a diverse dataset of URLs labeled as phishing or legitimate.
- Preprocessing the data to extract relevant features like URL length, domain age, presence of suspicious keywords, etc.

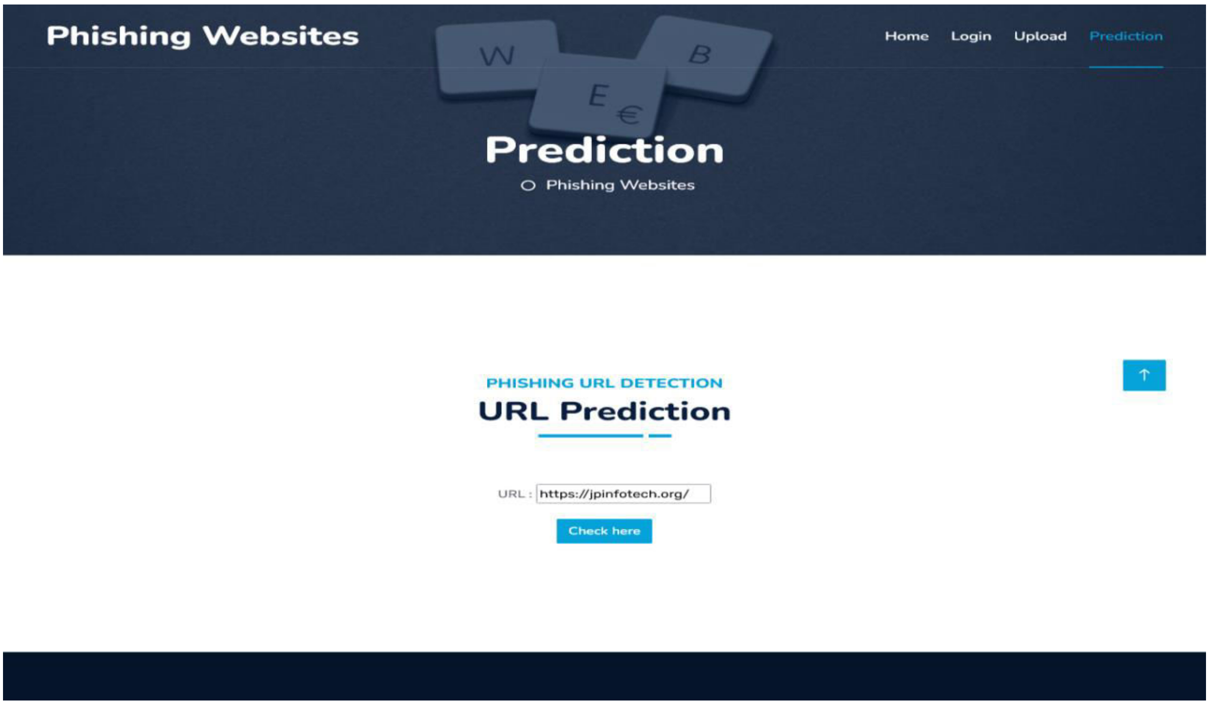
Feature Selection and Engineering:

- Choosing appropriate features that can effectively distinguish between phishing and legitimate websites.
- Engineering new features that might enhance the model's performance.

Outcome of the Detection:

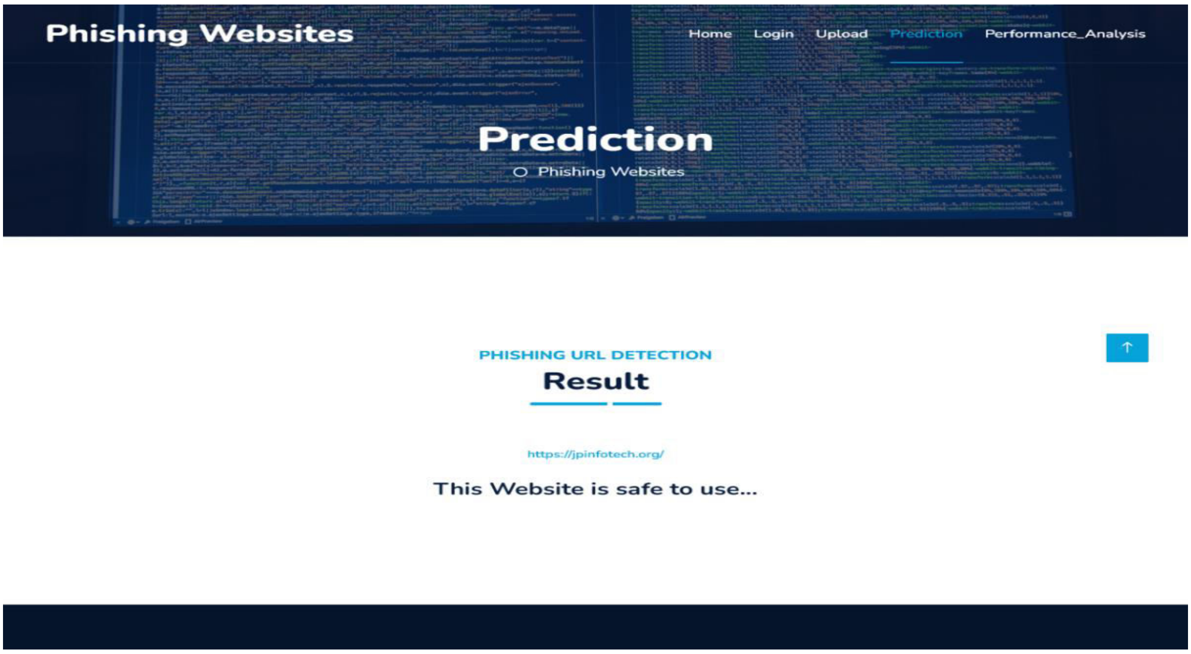
- The primary outcome is a trained machine learning model capable of classifying new, unseen URLs as phishing or legitimate with a certain level of accuracy.
- This model can be used in real-time systems to automatically flag potential phishing websites.

Snapshots:



URL : https://jpinfotech.org/

Figure 1: URL Prediction



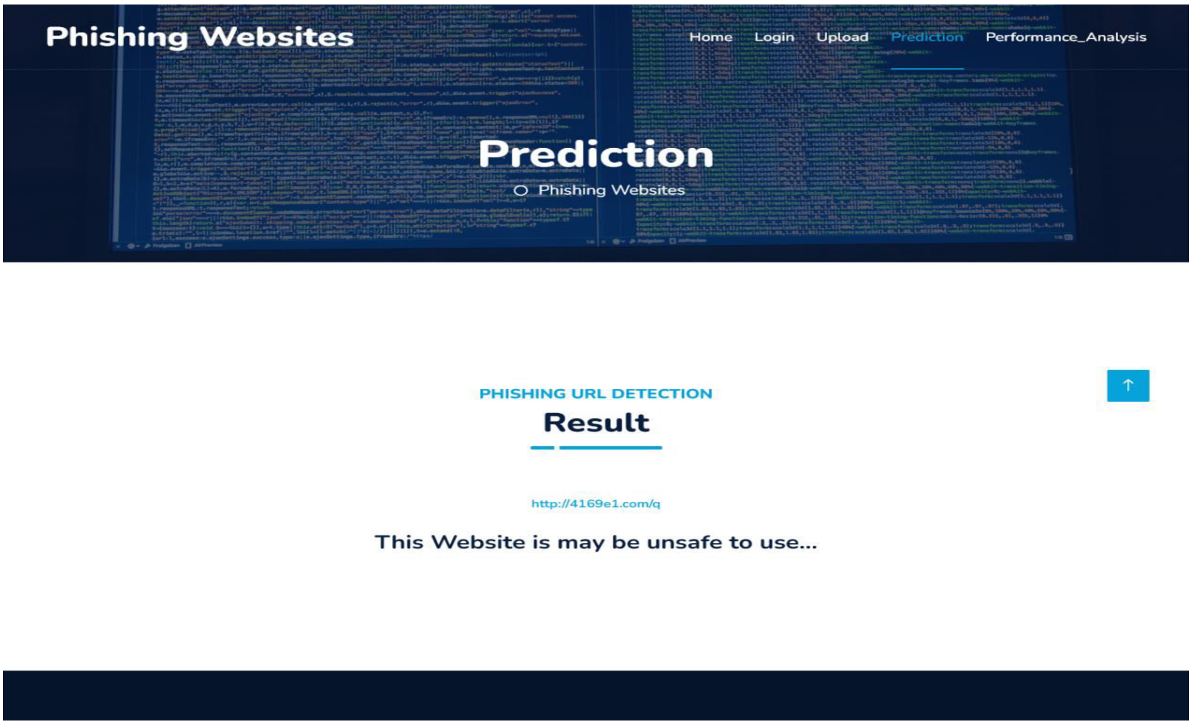
PHISHING URL DETECTION

Result

https://jpinfotech.org/

This Website is safe to use...

Figure 2: Result is safe to use



PHISHING URL DETECTION

Result

http://4169e1.com/q

This Website is may be unsafe to use...

Figure 3: Result is unsafe to use

V. CONCLUSION

It is remarkable that an effective anti-malware system can anticipate malware attacks within a reasonable amount of time. accepting that expanding the scope of malware site detection necessitates having a good anti-malware device accessible at a reasonable time. Gradient Boosting Classifier is all that is used in the current system to identify malware websites. With the lowest rate of false positives, we used a Gradient Boosting Classifier to achieve a detection accuracy of 97%.

REFERENCES

[1] Steve Sheng, Brad Wardman, Gary Warner, Jason Hong, Lorrie Faith Cranor, and Chengshan Zhang In CEAS 2009: Proceedings of the 6th Conference on Email and Anti-Spam, Mountain View, California, USA, July 16-17, 2009, Malware Blacklists: An Empirical Study

[2] A Literature Review on Malware Detection 2091-2121 in IEEE Communications Surveys and Tutorials, vol. 15, no. 2, by Andrew Jones, Mahmoud Khonji, and Youssef Iraqi, Senior Member 4, 2013. 2013.

[3] Many Understanding and Assisting Users' Online Choices with Nudges for Privacy and Security, Article No. 50(3), Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, and Florian Schaub 44, 2017 ACM Computing Surveys

[4] Mara M. and Helena Matute I'm looking for phishers. Moreno-Fernández, Fernando Blanco, Pablo Garaizar The sensitivity of Internet users to visual deception indicators needs to be increased in order to combat electronic fraud. pp. 421-436 in Computers in Human Behavior, Vol.69, 2017.

[5] F.J. Overink, M. L. Junger Montoya. Priming and warnings do not work to stop social engineering attacks. pp. 75-87 in Computers in Human Behavior, Vol. 66, 2017. 2017.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com